

Vers une base de données raisonnée des modélisations graphiques pour étudier le langage et les langues

Nicolas Mazziotta, Jacques François, Sylvain Kahane

DANS **TRAVAUX DE LINGUISTIQUE** 2023/2 (N° 87), PAGES 147 À 155
ÉDITIONS **DE BOECK SUPÉRIEUR**

ISSN 0082-6049

ISBN 9782807380318

DOI 10.3917/tl.087.0147

Article disponible en ligne à l'adresse

<https://www.cairn.info/revue-travaux-de-linguistique-2023-2-page-147.htm>



CAIRN.INFO
MATIÈRES À RÉFLEXION

Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...

Flashez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.



Distribution électronique Cairn.info pour De Boeck Supérieur.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

VERS UNE BASE DE DONNÉES RAISONNÉE DES MODÉLISATIONS GRAPHIQUES POUR ÉTUDIER LE LANGAGE ET LES LANGUES

Nicolas MAZZIOTTA, Jacques FRANÇOIS et Sylvain KAHANE*

Dans ce court article conclusif, nous voudrions proposer des pistes de réflexion pour l'élaboration d'une base de données raisonnée des diagrammes en linguistique – le terme *diagramme* est entendu ici dans le sens général peircien d'« icône de relations » présenté dans l'introduction. Une telle base de données se révélera utile en premier lieu pour éclairer les linguistes confirmés autant que les doctorants sur leurs pratiques : concentrer en un endroit un grand nombre de représentations concurrentes parmi lesquelles choisir de manière raisonnée a un intérêt méthodologique majeur. D'autre part, la base de données nourrira des expériences didactiques sur la pertinence des diagrammes en contexte d'apprentissage (Bélanger et Gauvin, 2020 ; Winn, 1991, 1993). En dehors de ces aspects que l'on pourrait qualifier d'utilitaires, ce répertoire fournira les matériaux nécessaires à des études sémiotiques et rhétoriques (Badir, 2008 ; Klinkenberg, 2010) qui nous éclaireront davantage sur le fonctionnement des diagrammes. Enfin, ces diagrammes et leur développement intéressent l'histoire des théories linguistiques au même titre que les productions textuelles des auteurs. On peut parfois montrer l'existence d'influences au sein même des diagrammes employés (Brittain, 1973 ; Coseriu, 1980 ; Mazziotta, 2016, 2020).

Si l'intérêt de pareille base de données ne fait aucun doute, une réflexion reste à mener sur les fondements conceptuels de sa mise en place, ce que nous proposons de faire ici. Dans un premier temps, nous rappelons la diversité des modélisations graphiques pour étudier le langage et les langues, tant du point de vue de leur constitution que du point de vue des usages que nous en faisons (1). Dans un deuxième temps, nous présentons

* Nicolas Mazziotta, Université de Liège, Traverses, nicolas.mazziotta@uliege.be ; Jacques François, Université de Caen-Normandie, CRISCO, E.A. 4255, jfrancois@interlingua.fr ; Sylvain Kahane, Université Paris-Nanterre et Institut Universitaire de France, MoDyCo, U.M.R. 7114, skahane@parisnanterre.fr.

brèvement quatre classifications différentes (2). La dernière section présente la structure conceptuelle que nous proposons de manière non technique (3).

1. Diversité des objets, diversité des usages

L'introduction du présent volume met en quelque sorte en évidence l'*invariant* qui en fonde la pertinence : les inscriptions graphiques des analyses linguistiques permettent de manipuler la connaissance représentée et d'en découvrir de nouvelles en effectuant des opérations concrètes sur les entités graphiques. Malgré cet invariant, les contributions du présent volume démontrent la grande diversité des modélisations diagrammatiques en linguistique, mais également la grande diversité des emplois des diagrammes. En se référant aux articles qui précèdent, qui n'illustrent qu'une petite partie de cette diversité, on peut synthétiser comme suit ces deux ordres de différences.

La diversité des modélisations tient tout d'abord au domaine de la linguistique qui est concerné. Chaque domaine définit ses problématiques et propose des diagrammes spécifiques qui inscrivent des structures conceptuelles différentes :

1. La linguistique comparée propose notamment des diagrammes arborescents phylogénétiques pour représenter la généalogie des langues d'une même famille. L'article « Les styles collectifs de pensée visuelle en sciences du langage » (Jacques François) a montré que diverses formes concurrentes étaient mobilisées, dont certaines aptes à rendre compte de la chronologie (« glottochronologie ») sous-tendant l'arborescence.
2. La sémantique typologique propose différentes sortes de cartes sémantiques pour évaluer la construction des sémèmes. L'article « Les cartes sémantiques en typologie des langues » (Sémir Badir et Stéphane Polis) montre qu'il y a différents types de cartes sémantiques, avec notamment d'importants contrastes entre les cartes synchroniques / contrastives et les cartes diachroniques.
3. La contribution « Les diagrammes en sémantique cognitive » (Gilles Col) a discuté des diagrammes pour illustrer et pour représenter les rapports cognitifs. Elle a notamment plaidé pour que soient élaborées de nouvelles inscriptions plus dynamiques, qui ajoutent aux réseaux une dimension temporelle.
4. La dialectologie représente la distribution diatopique de formes liées dans un rapport sémasiologique ou, plus souvent, onomasiologique, sur une carte géolinguistique. L'article « Ce que contiennent (et ce que ne contiennent pas) les cartes géolinguistiques » (Esther Baiwir)

- a exploré la question du niveau d'élaboration des cartes linguistiques par rapport aux matériaux dont elles inscrivent la répartition.
5. Les diagrammes syntaxiques représentent les rapports syntagmatiques au sein d'une phrase ou d'un énoncé sous la forme d'un arbre ou d'un graphe. L'article « Grammaires graphiques » (Sylvain Kahane et Nicolas Mazziotta) a montré qu'ils étaient manipulables d'un point de vue formel au même titre que toute structure mathématique, c'est-à-dire comme support d'opérations. L'article « Délimitation des diagrammes syntaxiques dans la grammaire scolaire » (Nicolas Gregov) a quant à lui posé la question de l'utilisation didactique de ce type de diagramme.

Les articles de ce volume démontrent clairement que la délimitation du domaine ne suffit pas à déterminer à elle seule les propriétés des diagrammes employés. Il existe une importante variation au sein même des spécialités de la linguistique, ainsi qu'une démarche de recherche vivante (bien que parfois discrète) dont le principal objectif est de faire évoluer les inscriptions graphiques en vue d'améliorer leur adéquation à nos préoccupations scientifiques.

À cela s'ajoute une importante liberté par rapport aux formes de la représentation d'un même contenu. De manière générale, un objet aussi « simple » qu'une structure arborescente peut être représenté sous onze formes prototypiques différentes¹ selon Lima (2014). Or, comme l'a fait remarquer la littérature sur les diagrammes, il existe des manières alternatives de représenter les mêmes contenus et ces manières alternatives n'ont pas les mêmes potentialités en termes cognitifs (Shimojima, 2015). À cet égard, il en est pour la linguistique comme pour de nombreux autres domaines du savoir : choisir la bonne représentation est un enjeu méthodologique capital – on pense p. ex. aux statistiques ou aux inscriptions des structures moléculaires et atomiques en chimie et en physique, dont la sélection dépend de l'application².

2. Classifications existantes

La question de la distinction entre les objets graphiques diagrammatiques et les représentations picturales (*depictions*) demeure un sujet de discussion (Norman, 2003), mais nous la considérerons comme intuitivement résolue pour nous concentrer sur les diagrammes tels qu'ils ont été définis dans l'introduction et examinés dans les différentes contributions. On peut tenter de classer les représentations graphiques en général. Un exemple classique est la proposition de Jacques Bertin dans sa *Sémiologie graphique : les diagrammes, les réseaux, les cartes* (Bertin, 2005 [1967]). Le titre de l'ouvrage révèle la sélection (et, incidemment, la définition) de trois classes

d'inscriptions graphiques, définies et décrites selon une approche et une terminologie spécifiques à l'auteur. L'idiosyncrasie du système d'analyse nous alerte d'emblée sur la difficulté que nous allons rencontrer, même en nous centrant exclusivement sur les diagrammes des linguistes. Établir une typologie raisonnée n'est pas chose aisée ; l'incompatibilité des différentes approches de la question semble inévitable.

En matière d'inventaire raisonné des diagrammes linguistiques, il nous semble que les quatre propositions suivantes méritent d'être mentionnées.

1. *Graphic representations of models in linguistic theory* (Stewart, 1976). – La première partie de l'ouvrage de Stewart distingue les arbres, les matrices / grilles et les boîtes en utilisant des critères liés à la structure formelle des diagrammes et à la valeur abstraite associée à chacun de ces types. Ces derniers sont ensuite détaillés en fonction des analyses qui peuvent y être inscrites : phylogénie, taxonomie, analyse componentielle et analyse en constituants pour les arbres, analyse componentielle exclusivement pour les matrices, analyse componentielle, analyse en constituants et amalgamations pour les boîtes (voir les détails de la classification dans Stewart, 1976 : 6-75 et en particulier la synthèse, 74).
2. *Towards a Typology of Diagrams in Linguistics* (Smessaert et Demey, 2018). – La proposition de Smessaert et Demey insiste sur la distinction entre les propriétés sémiotiques des diagrammes et la matière relative à l'épistémologie de la linguistique qu'ils inscrivent. Ils posent ainsi six paramètres de description dichotomiques proprement linguistiques : 1° unilingue / multilingue ; 2° relatif à une information synchronique / diachronique ; 3° centré sur un domaine de la linguistique ou sur plusieurs (phonétique, morphologie, syntaxe, etc.) ; 4° relatif à une information sur les unités linguistiques observées ou sur l'analyse (p. ex. un arbre génétique classe les langues et non les unités linguistiques) ; 5° qualitatif / quantitatif ; 6° proprement linguistique / interdisciplinaire (neurolinguistique, sociolinguistique, etc.). Deux paramètres sémiotiques, applicables aux diagrammes en général, s'ajoutent à cette liste : 1° iconique / symbolique ; 2° dynamique (représentant une séquence ordonnée ou une succession) / statique (ne représentant pas ce type de contenu). Les auteurs appliquent ensuite leur analyse à un échantillon de visualisations graphiques pour en fournir une analyse componentielle, qui caractérise les diagrammes par rapport à toutes ces dichotomies.
3. *Visuelle Linguistik* (Bubenhofer, 2020). – Après une introduction qui délimite la notion de *diagramme* dans le champ des outils scientifiques visuels, le récent livre de Bubenhofer propose une classification en

cinq « figures diagrammatiques de base » sur la base de leur logique de construction, associée à l'histoire de leur constitution : les listes, les cartes (géographiques ou non), les tableaux (*Partituren*, lit. “partitions”), les vecteurs (après conversion des données en listes de nombres) et les graphes (arbres et réseaux). L'auteur accorde une grande importance à l'utilisation des diagrammes (chap. 6-8), en particulier aux transformations d'un type de diagramme à un autre et au potentiel pratique de chacun de ces types. Il discute la possibilité d'altérer les diagrammes de manière dynamique (notamment à l'aide d'interfaces informatiques) et la tension entre les approches inductive et hypothético-déductive.

4. *Les techniques de visualisation dans les sciences du langage* (François, 2022). – Il s'agit d'un répertoire annoté de *visualisations* (terme considéré comme englobant et sous-spécifié). Le répertoire est précédé d'une typologie de ces représentations. Celle-ci se fonde sur les définitions du TLF et le dictionnaire électronique des synonymes du CRISCO³ pour distinguer les signifiés des termes généraux *carte*, *diagramme*, *graphique*, *schéma* (voir François, 2022 : chap. 2 pour l'analyse et les distinctions entre ces types). L'auteur propose une classification permettant de caractériser les visualisations comme relevant de types mixtes plus ou moins utilisés (schéma-diagramme, graphique-carte, etc.). Chaque diagramme du répertoire est associé à une fiche descriptive qui décrit ce que le diagramme signifie au sein du texte dont il est extrait.

Ces quatre propositions de classification nous montrent qu'il n'y a pas de consensus en matière d'analyse et de taxinomie des représentations visuelles en linguistique. À la grande variété des diagrammes et de leurs utilisations vient se superposer la diversité de leur description.

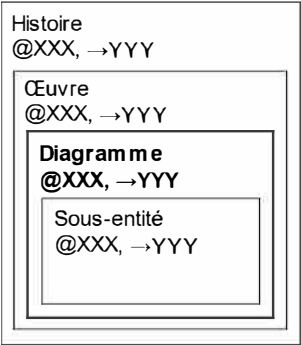
3. Proposition d'architecture fondamentale

Dans cette section, nous décrivons de manière non technique une proposition d'architecture pour les fiches de la base de données. De ce qui précède, nous retiendrons cinq aspects qui nous paraissent importants. Du point de vue de leur description intrinsèque, les diagrammes sont des objets *sémio-tiques* dont les propriétés peuvent être décrites en termes relevant de ce domaine du savoir. En tant que signes iconiques, les diagrammes ont un signifiant articulé à un signifié relevant d'une ou plusieurs facettes de la linguistique. Leur caractère articulé implique la présence de plusieurs entités graphiques (concepts réifiés, cf. notre introduction) organisées dans l'espace bidimensionnel. Du point de vue de leur ancrage, les modélisations visuelles sont des composantes d'*énoncés plus larges* (œuvres

ou ensembles d’œuvres) qui font intervenir d’autres inscriptions de la connaissance (discours en langue naturelle, autres diagrammes, etc.). Les représentations peuvent être qualifiées de *pratiques disciplinaires historiques*. Eu égard à ces deux derniers aspects, il peut exister des rapports manifestes, synchroniques et diachroniques, *entre les diagrammes et entre leurs sous-entités* (conversions de types, simplifications, héritage de pratiques, modifications de traditions, etc.).

Ces aspects impliquent que chaque diagramme est situé au cœur de rapports méréologiques hiérarchisés : chaque ensemble de *pratiques historiques* (écoles, tranche chronologique, etc.) comprend un certain nombre d’œuvres qui comprennent elles-mêmes des *diagrammes* constitués de leurs *sous-entités* (Figure 1) comme fondement de la base de données. Chacun de ces quatre éléments est susceptible d’être annoté, d’une part par une description de ses attributs intrinsèques associant un type à une valeur (précédés de « @ » dans la Figure 1), d’autre part des déterminations extrinsèques, issues de la mise en relation avec un autre élément de la base (précédées de « → » dans la Figure 2).

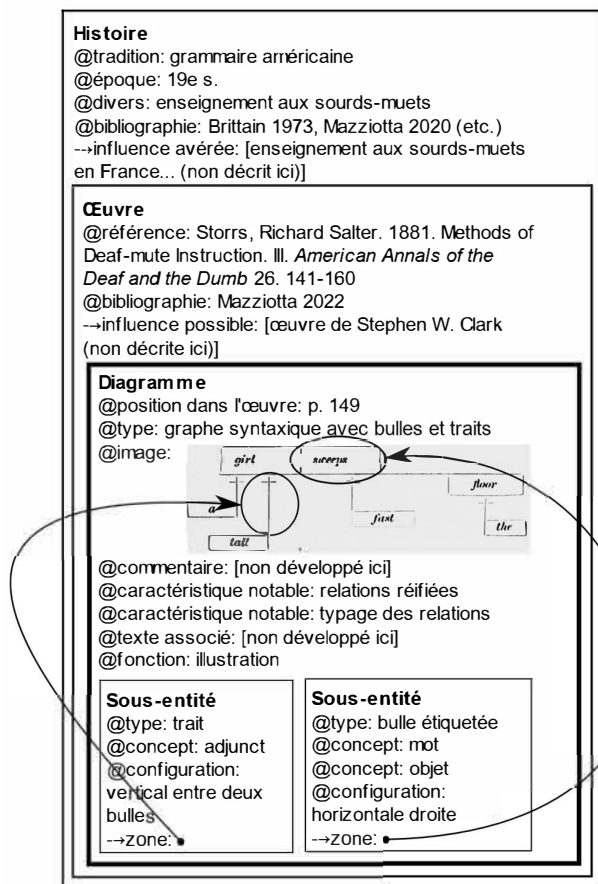
Figure 1 : Imbrication des éléments fondamentaux de la base de données. Le diagramme, central, est mis en évidence en gras. (plusieurs éléments du même type peuvent apparaître à chaque niveau)



Pour permettre les extractions, les attributs soit font partie d’une liste définie *a priori* (ontologie), soit sont définis de manière souple. De la même manière, les rapports de détermination peuvent relever de types prédéfinis ou demeurer versatiles.

À titre d’exemple, la Figure 1 représente la fiche d’un diagramme syntaxique proposé par un grammairien américain nommé Richard Salter Storrs, qui enseignait notamment la grammaire anglaise à des sourds-muets (Mazziotta, 2022) – la description proposée ici n’a qu’un but illustratif et pourrait être critiquée et amplifiée.

Figure 2 : Exemple de traitement partiel de diagramme dans la base de données (les relations internes à l'exemple sont représentées par des flèches)



Nous insistons sur la flexibilité que devra conserver la base, de façon à ce que chaque chercheur puisse déterminer quels attributs et quelles relations sont pertinents pour ses questions de recherche. Ainsi le type de diagramme « arbre à bulle » pourrait être considéré comme trop spécifique par un autre chercheur, qui pourrait vouloir préciser que les bulles sont en réalité des boîtes anguleuses ou employer une autre terminologie ne mobilisant pas le concept de *réification*, ou encore supprimer certains attributs et certaines relations. Seule la structure fondamentale, constituée de l'imbriication méréologique *histoire* / *œuvre* / *diagramme* / *sous-entité*, ainsi que de la présence d'attributs et de relations typés, devra être considérée comme un invariant.

Ces premières propositions rendent possible une réflexion sur la mise en œuvre concrète d'une base de données qui permettra d'agréger les analyses des chercheurs en fonction de leurs besoins.

NOTES

1. Nous renvoyons à son ouvrage, dont chaque chapitre correspond à une de ces formes : arbres figuratifs, verticaux, horizontaux, multidirectionnels, radiaux et hyperboliques, ainsi que différentes *treemaps* – rectangulaires, de Voronoï, circulaires, de type *sunburst* ou *icicle*.

2. Par exemple, en statistiques exploratoires, la visualisation d'un espace issu de la décomposition des données se fait traditionnellement à l'aide de flèches superposées sur le plan en cas d'analyse en composantes principales, mais pas en cas d'analyse factorielle des correspondances (voir Lebart, Morineau et Piron, 1998 pour une description technique de ces méthodes).

3. <https://crisco2.unicaen.fr/des/>.

RÉFÉRENCES BIBLIOGRAPHIQUES

- BADIR S., 2008, « Sémiotique des graphiques / Graphiques de sémiotique », *Visible*, 4, p. 11-59.
- BÉLANGER V. et GAUVIN I., 2020, « Arbres syntaxiques et enseignement inductif au service de l'enseignement de la grammaire. Le cas de la notion "sujet de phrase" en première secondaire », in TREMBLAY O., FALARDEAU É., BOYER P. et GAUVIN I., *Diffusion et influences des recherches en didactique du français*, Namur, Presses Universitaires de Namur, p. 143-80.
- BERTIN J., 2005⁴ [1967], *Sémiologie graphique: les diagrammes, les réseaux, les cartes*, 4^e éd., Paris, EHESS.
- BRITAIN R. C., 1973, *A critical history of systems of sentence diagramming in English*, University of Texas at Austin, <https://books.google.se/books?id=6Ve8pwAACAAJ>.
- BUBENHOFER N., 2020, *Visuelle Linguistik: zur Genese, Funktion und Kategorisierung von Diagrammen in der Sprachwissenschaft*, Berlin, De Gruyter.
- COSERIU E., 1980, « Un précurseur méconnu de la syntaxe structurale: H. Tiktin », in BINGEN J., COUPEZ A. et MAWET F., *Recherches de linguistique: hommages à Maurice Leroy*, Bruxelles, Éditions de l'Université de Bruxelles, p. 48-62.
- FRANÇOIS J., 2022, « Les techniques de visualisation dans les sciences du langage », <https://normandie-univ.hal.science/hal-03797302/>.
- KLINKENBERG J.-M., 2010, « À quoi servent les schémas ? Tabularité et dynamisme linéaire », *Protée*, 37, 3, <https://doi.org/10.7202/038806ar>, p. 65-73.
- LEBART L., MORINEAU A. et PIRON M., 1998², *Statistique exploratoire multidimensionnelle*, Sciences sup., Paris, Dunod.
- LIMA M., 2014, *The book of trees: visualizing branches of knowledge*, New York, Princeton, Architectural Press.

- MAZZIOTTA N., 2016, « Drawing syntax before syntactic trees: Stephen Watkins Clark's sentence diagrams (1847) », *Historiographia Linguistica*, 3, <https://doi.org/10.1075/hl.43.3.03maz>, p. 301-342.
- MAZZIOTTA N., 2020, « Grammar and graphical semiotics in early syntactic diagrams: Clark (1847) and Reed-Kellogg (1876) », in FORTIS J.-M. et AUSSANT E., *Historical Journey in a Linguistic Archipelago: Descriptive Concepts and Case Studies*, Berlin, Language Science Press, p. 67-81, <https://doi.org/10.5281/ZENODO.4269413>.
- MAZZIOTTA N., 2022, « Through the eyes of grammar: Richard Salter Storr's (1830-1884) sentence-maps », in DENECKER T., DESMET P., JOOKEN L., LAUWERS P., VAN HAL T. et VAN ROOY R., *The architecture of grammar: Studies in linguistic historiography in honor of Pierre Swiggers*, Leuven, Peeters, p. 363-389.
- NORMAN J., 2003, « Differentiating diagrams: a new approach », in ANDERSON M., CHENG P. et HAARSLEV V., *Theory and Application of Diagrams. Diagrams 2000*, Berlin, Heidelberg, Springer Berlin Heidelberg, p. 105-116, https://doi.org/10.1007/3-540-44590-0_13.
- SHIMOJIMA A., 2015, *Semantic properties of diagrams and their cognitive potentials*, Stanford (California), CSLI Publications.
- SMESSAERT H. et DEMEY L., 2018, « Towards a typology of diagrams in linguistics », in CHAPMAN O., STAPLETON G., MOKTEFI A., PEREZ-KRIZ S. et BELLUCCI F., *Diagrams 2018. Diagrammatic representation and inference*, p. 236-244.
- STEWART A. H., 1976, *Graphic representation of models in linguistic theory*, Bloomington, Indiana University Press.
- WINN W., 1991, « Learning from maps and diagrams », *Educational Psychology Review*, 3, 3, <https://doi.org/10.1007/BF01320077>, p. 211-47.
- WINN W., 1993, « An account of how readers search for information in diagrams », *Contemporary Educational Psychology*, 18, 2, <https://doi.org/10.1006/ceps.1993.1016>, p. 162-185.